Learning Conditional Distributions using Mixtures of Truncated Basis Functions

Inmaculada Pérez-Bernabé[1] Antonio Salmerón[1] Helge Langseth[2]

[1]Dept. Mathematics, University of Almería, Spain

[2]Dept. Computer and Information Science. Norwegian University of Science and Technology, Trondheim, Norway

- MoTBFs provide a flexible framework for hybrid BNs.
- Accurate approximation of known models.
- Learning from data.

Previous models in this area

- Conditional Linear Gaussian model (CLG) (Lauritzen (1992)).

Previous models in this area

- Conditional Linear Gaussian model (CLG) (Lauritzen (1992)).
- Mixtures of Truncated Exponentials (MTEs) (Moral et al. (2001)).

Previous models in this area

- Conditional Linear Gaussian model (CLG) (Lauritzen (1992)).
- Mixtures of Truncated Exponentials (MTEs) (Moral et al. (2001)).
- Mixtures of Polynomials (MoPs) (Shenoy and West, (2011)).

Current approach for learning MoTBFs from dat

The MoTBF framework is based on the abstract notion of real-valued *basis functions* $\psi(\cdot)$, which include both polynomial and exponential functions as special cases.

Current approach for learning MoTBFs from dat

The MoTBF framework is based on the abstract notion of real-valued *basis functions* $\psi(\cdot)$, which include both polynomial and exponential functions as special cases.

MoTBF Potential

$$f(x) = \sum_{i=0}^{k} c_i \, \psi_i(x)$$

The MoTBF framework is based on the abstract notion of real-valued *basis functions* $\psi(\cdot)$, which include both polynomial and exponential functions as special cases.

MoTBF Potential

$$f(x) = \sum_{i=0}^{k} c_i \, \psi_i(x)$$

MoTBF Density

$$\int_{\Omega_X} f(x) \, \mathrm{d}x = 1$$

Univariate case. MoPs

- We use the method in (Langseth et al. 2014).
- Given a sample $D = \{x_1, \ldots, x_N\}$, construct the empirical CDF:

$$G_N(x) = \frac{1}{N} \sum_{\ell=1}^{N} \mathbf{1}\{x_\ell \leq x\}, \quad x \in \mathbb{R},$$

  where $\mathbf{1}\{\cdot\}$ is the indicator function.
- Then we fit a potential whose derivative is an MoTBF, to the empirical CDF using least squares.
- Though this is not properly ML, we have shown in (Langseth et al. 2014) that it is competitive in terms of likelihood and numerically more stable.

Univariate case. MoPs

- As an example, if we use polynomials as basis functions, $\Psi = \{1, x, x^2, x^3, \ldots\}$, the parameters can be obtained solving the optimization problem

$$\begin{aligned} \textbf{minimize} \quad & \sum_{\ell=1}^{N} \left(G_N(x_\ell) - \sum_{i=0}^{k} c_i \, x_\ell^i \right)^2 \\ \textbf{subject to} \quad & \sum_{i=1}^{k} i \, c_i \, x^{i-1} \geq 0 \quad \forall x \in \Omega, \\ & \sum_{i=0}^{k} c_i \, a^i = 0 \text{ and } \sum_{i=0}^{k} c_i \, b^i = 1, \end{aligned} \tag{1}$$

- We use `solvQP` from R package `quadprog`.

Estimation of univariate MoPs

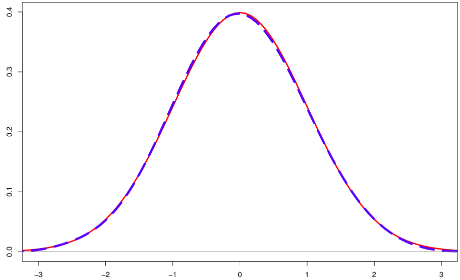

Figure : A standard normal density (solid line) overlaid to an MoTBF approximation (dashed line) restricted to interval [-3,3].

Multivariate case. MoPs

- We have $D = \{(x_1, y_1), \dots, (x_N, y_N)\}$ and

$$G_N(\mathbf{x}) = \frac{1}{N}\sum_{\ell=1}^{N} \mathbf{1}\{\mathbf{x}_\ell \leq \mathbf{x}\}, \quad \mathbf{x} \in \Omega_{\mathbf{X}} \subset \mathbb{R}^d.$$

- The optimization problem to solve is

$$\begin{aligned} \textbf{minimize} \quad & \sum_{\ell=1}^{N} \left(G_N(\mathbf{x}_\ell) - F(\mathbf{x}_\ell)\right)^2 \\ \textbf{subject to} \quad & \frac{\partial^d F(\mathbf{x})}{\partial x_1, \dots, \partial x_d} \geq 0 \quad \forall \mathbf{x} \in \Omega_{\mathbf{X}}, \\ & F\left(\Omega_{\mathbf{X}}^-\right) = 0 \text{ and } F\left(\Omega_{\mathbf{X}}^+\right) = 1. \end{aligned} \tag{2}$$

where $F(\mathbf{x}) = \sum_{\ell_1=0}^{k} \cdots \sum_{\ell_d=0}^{k} c_{\ell_1,\ell_2,\dots,\ell_d} \prod_{i=1}^{d} x_i^{\ell_i}$,

Estimation of bivariate MoPs

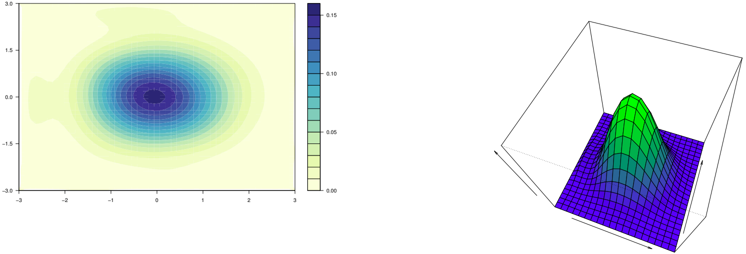

Figure : The contour and the perspective plots of the result of learning a MoP from $N = 1000$ samples drawn from bivariate standard normal distributions with $\rho = 0$.

Estimation of bivariate MoPs

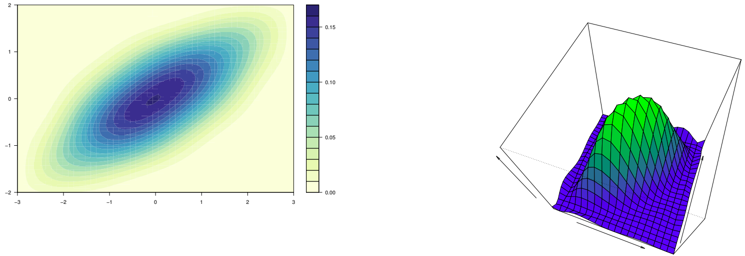

Figure : The contour and the perspective plots of the result of learning a MoP from $N = 1000$ samples drawn from bivariate standard normal distributions with $\rho = 0.99$.

Conditional MoPs

Using the minimization program in Equation 2 and by the definition of a conditional probability density we will have:

$$f(x|\mathbf{z}) \leftarrow \frac{f(x,\mathbf{z})}{f(\mathbf{z})}$$

MoPs are not closed under division, thus $f(x|\mathbf{z})$ will not lead to a legal MoP-representation of a conditional density.

Conditional MoPs

An alternative was previously proposed, where the influence that the parents $\mathbf{Z}$ have on X was encoded only through the partitioning of the domain of $\mathbf{Z}$ into hyper-cubes.

Conditional MoPs

An alternative was previously proposed, where the influence that the parents $\mathbf{Z}$ have on X was encoded only through the partitioning of the domain of $\mathbf{Z}$ into hyper-cubes.

- H. Langseth, T.D. Nielsen, I. Pérez-Bernabé, A. Salmerón (2014) Learning mixtures of truncated basis functions from data. International Journal of Approximate Reasoning 55, 940-956.

Conditional MoPs

- Compute an MoP representation for $f(x, \mathbf{z})$ using the program in Equation 2
- *Calculate* $f(\mathbf{z}) = \int_{\Omega_x} f(x, \mathbf{z}) \mathrm{d}x$.
- The conditional distribution defined through Equation 3 is our target, leading to the following optimization program

$$
\begin{aligned}
\textbf{minimize} \quad & \sum_{\ell=1}^{N} \left(\frac{f(x_\ell, \mathbf{z}_\ell)}{f(\mathbf{z}_\ell)} - f(x_\ell | \mathbf{z}_\ell) \right)^2 \\
\textbf{subject to} \quad & f(x|\mathbf{z}) \geq 0 \quad \forall (x, \mathbf{z}) \in (\Omega_X \times \Omega_{\mathbf{z}}) .
\end{aligned} \tag{3}
$$

- Normalizing the distribution the solution of this problem.

Two different scenarios:

- $Y \sim \mathcal{N}(\mu = 0, \sigma = 1)$ and $X | \{Y = y\} \sim \mathcal{N}(\mu = y, \sigma = 1)$.
- $Y \sim \text{Gamma}(\text{rate} = 10,\ \text{shape} = 10)$ and $X | \{Y = y\} \sim \text{Exp}(\text{rate} = y)$.

For each scenario, we generated 10 data-sets of samples $\{X_i, Y_i\}_{i=1}^{N}$, where the size is chosen as $N = 25, 500, 2500, 5000$.

Mean square error

N	$f_{X\|Y}(x\|y)$	*Split Method*	*MoTBF Algorithm*	*B-Splines Method*
25	y=-0.6748	0.1276	0.0848	**0.0103**
	y=0.00	0.1254	0.0936	**0.0089**
	y=0.6748	0.1279	0.1416	**0.0105**
500	y=-0.6748	0.0256	0.0453	**0.0025**
	y=0.00	0.0317	0.0117	**0.0009**
	y=0.6748	0.0246	0.0411	**0.0020**
2500	y=-0.6748	0.0031	0.0019	**0.0006**
	y=0.00	0.0064	0.0010	**0.0002**
	y=0.6748	0.0058	0.0024	**0.0006**
5000	y=-0.6748	0.0019	0.0018	**0.0006**
	y=0.00	0.0074	0.0009	**0.0002**
	y=0.6748	0.0019	0.0020	**0.0006**

Table : Average MSE between the different methods to obtain MoP approximations and the true conditional densities for each set of 10 samples, where $Y \sim \mathcal{N}(0,1)$ and $X|Y \sim \mathcal{N}(y,1)$.

Estimation of conditional MoPs

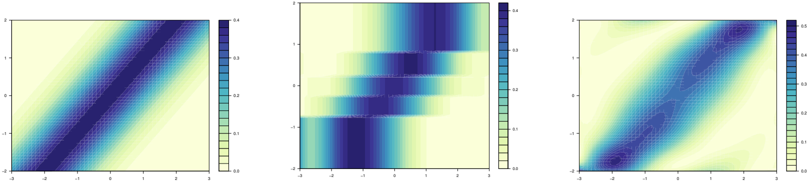

Figure : True conditional density, the MoP produced by the method introduced in Langseth et al. 2014 and the MoP obtained by the new proposal.

Mean square error

N	$f_{X\|Y}(x\|y)$	*Split Method*	*MoTBF Algorithm*	*B-Splines Method*
25	y=0.7706	0.4054	**0.0083**	0.0131
	y=0.9684	0.4703	**0.0081**	0.0225
	y=1.1916	0.5473	**0.0229**	0.0374
500	y=0.7706	0.0158	0.0037	**0.0012**
	y=0.9684	0.0048	0.0034	**0.0022**
	y=1.1916	0.0118	**0.0039**	0.0057
2500	y=0.7706	0.0064	**0.0025**	**0.0025**
	y=0.9684	0.0080	**0.0024**	0.0043
	y=1.1916	0.0029	**0.0046**	0.0074
5000	y=0.7706	**0.0013**	0.0021	0.0015
	y=0.9684	0.0091	**0.0015**	0.0022
	y=1.1916	**0.0026**	0.0029	0.0032

Table : Average MSE between the different methods to obtain MoP approximations and the true conditional densities for each set of 10 samples, where $Y \sim Gamma(rate = 10, shape = 10)$ and $X|Y \sim Exp(y)$.

Estimation of conditional MoPs

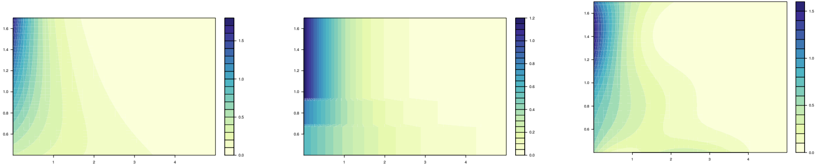

Figure : True conditional density, the MoP produced by the method introduced in Langseth et al. 2014 and the MoP obtained by the new proposal.

Conclusions

- We have developed a method for learning conditional MoTBFs.
- The advantage of this proposal with respect to the B-spline is that there is no need to split the domain of any variable.
- The experimental analysis suggests that our proposal is competitive with the B-spline approach in a range of commonly used distributions.
- We have done the appropriate implementation in R (R Development Core Team).

This research has been partly funded by the Spanish Ministry of Economy and Competitiveness, through project TIN2013-46638-C3-1-P and by Junta de Andalucía through project P11-TIC-7821 and by ERDF funds. A part of this work was performed within the AMIDST project. AMIDST has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 619209.